

November 26, 2024

***E-Filed***

The Honorable Thomas S. Hixson  
United States District Court for the Northern District of California  
San Francisco Courthouse, Courtroom E – 15th Floor  
450 Golden Gate Avenue  
San Francisco, CA 94102

Re: *Kadrey et al. v. Meta Platforms, Inc.*; Case No. 3:23-cv-03417-VC-TSH

Dear Magistrate Judge Hixson:

Plaintiffs in the above-captioned action (“Plaintiffs”) and Defendant Meta Platforms, Inc. (“Meta”) jointly submit this letter brief regarding issues related to Meta’s Responses and Objections to Plaintiffs’ recent sets of Requests for Production (“RFPs”). The parties met and conferred on November 20 and 21, 2024 but were unable to reach a resolution.

## I. PLAINTIFFS' STATEMENT

Plaintiffs request that the Court correct the following RFP deficiencies.

### A. Impermissibly Narrow Definitions

**“Llama Models”:** Meta now treats Llama 4 as relevant, but it objects to the inclusion of Llama 5—another model currently under development—within Plaintiffs’ discovery requests. Llama 5 is relevant for the same reasons the Court already ruled that Llama 4 is relevant. Dkt. 279 at 4.<sup>1</sup> Meta argues construing Plaintiffs’ discovery requests as encompassing Llama 5 would impose a perpetual obligation for Meta to supplement its document productions. That is no reason to shield relevant discovery—the class period runs “through the present,” Dkt. 133 at 14, and Meta rapidly develops new iterations of its Llama models. Courts require supplementation after the end of fact discovery where “material” changes to “project source code” will occur in “ongoing” intellectual property violations. *E.g., Valeo Schalter und Sensoren GmbH v. NVIDIA Corp.*, 2024 WL 3916101, at \*1-2 (N.D. Cal. Aug. 23, 2024). Meta need not perpetually rerun its search terms, but it must produce responsive Llama 5 documents that exist already and must continue to produce newly-created responsive material—including new source code.

**“Shadow Datasets”:** Plaintiffs use this term in RFP 81 to describe Meta’s use of pirated databases to train the Llama Models—“including but not limited to” six exemplar databases. Meta improperly limits its responses to just those six exemplars. Only Meta knows whether it used other pirated databases for training connection with its Llama models. If responsive materials exist for other databases, Meta should produce them.

**“Agreements”:** Meta improperly construes this term—which includes oral and inchoate agreements—to mean only written contracts. RFPs 91 and 130 ask for documents and communications regarding licensing agreements for the Llama models and AI training data. Existing documents reveal [REDACTED], yet Meta avoids responsive materials regarding those efforts by limiting Plaintiffs’ requests to written contracts (which don’t exist [REDACTED]). Meta should not be permitted to withhold such critical information.

### B. Wholesale Relevance Objections

**Licensing Voice Recordings:** RFP 114 requests documents related to Meta’s partnerships with celebrities whose voices are used to train Meta’s AI chatbot. This RFP is relevant to fair use. Meta asserts that it had no need to license Plaintiffs’ copyrighted works to train Llama, but to the extent Meta entered into licenses for use of voice recordings to train a similar AI platform, that decision plainly is relevant to Meta’s assertion that licensing wasn’t required in this case.

**Withheld EU Release:** RFP 115 involves Meta’s recent decision not to release its newest AI models to the European Union. This Request is relevant because Meta’s reticence to violate EU data privacy regulations, including a forthcoming requirement to disclose training datasets,<sup>2</sup> bears on Meta’s willful infringement and knowledge of the risks of training Llama on pirated works.

**Nonidentification of Training Databases:** Meta has decided not to publicly disclose the databases it uses to train more recent Llama models, despite branding Llama as an “open-source” platform. RFP 117 seeks documents related to that decision. This request is relevant, as Meta likely

<sup>1</sup> Plaintiffs raised Llama 5 in Dkt. 267 at 10 n.9 but the Court addressed only Llama 4, Dkt. 279 at 4.

<sup>2</sup> <https://www.theverge.com/2024/7/18/24201041/meta-multimodal-llama-ai-model-launch-eu-regulations>

made this decision because the training databases contain pirated versions of copyrighted works.

### C. Improper “Sufficient to Show” Limitations

**Licensing Agreements:** RFP 91 requests all documents related to licensing agreements for the Llama models. This is critical: Meta decided *not* to pay for licensing copyrighted works to train Llama and instead used pirated versions without compensating copyright holders. The market for licensing books is also relevant to fair use. The Court should order Meta to produce *all* responsive materials, including all licensing agreements, whether inchoate or executed, both written and oral.

**Financial Data:** RFPs 94 and 100-101 request financial documents for the Llama program. Meta’s counsel represented that Meta’s “sufficient to show” limitation will result in production of a certain subset of financial statements only. But Meta intends to withhold other responsive financial documents, including those involving its broader Generative AI division (*e.g.*, presentations with financials associated with licensing AI to third parties and/or integrating AI into Meta’s products, financial models, and emails discussing these issues), without substantiating any claim of burden.

**Training Guidelines:** RFP 106 asks for all documents related to Meta’s guidelines for the treatment of copyrighted material in connection with the Llama models. This is highly relevant to Meta’s knowledge of and willfulness in infringing Plaintiffs’ copyrights. To the extent any guidelines exist within Meta governing the treatment of copyrighted material, they all should be produced, along with any related documents and communications.

**Fair Use Communications:** RFP 123 asks for all non-privileged communications with third parties regarding Plaintiffs’ allegations and Meta’s defenses, including fair use. This too is highly relevant and Meta has failed to substantiate any claim of undue burden. The Court should order Meta to respond to this RFP as written, including by producing all third-party communications concerning fair use in connection with the use of copyrighted materials for LLM models.

### D. Commits and Pull Requests

RFP 120 requests source code “commits” and “pull requests.” “Commits” are save files for coding projects that allow programmers to preserve “snapshots” to see what changes were made and when. “Pull requests” are proposals to merge coding changes, which also allow programmers to review and discuss proposed changes. This data is frequently discoverable, *e.g.*, *SmartLinx Solutions LLC v. Zeif*, 2023 WL 11964457, at \*3 (D.S.C. Aug. 31, 2023) (granting motion to compel pull request schedule). Meta’s counsel represents that the data is available in Meta’s on-site data room, but Meta refuses to produce the commits/pull requests as text files, meaning Plaintiffs’ experts can only view them on a computer within Meta’s on-site data room during times Meta allows them inside. That is not sufficient. Plaintiffs’ experts need the ability to run searches through the files to perform their analysis without having to physically be present at Meta’s offices.

### E. Requests Allowing Assessment of Privilege

RFP 125 requests documents sufficient to show the identity of persons included on listservs referenced within Meta’s privilege logs. A party claiming privilege must identify the senders and recipients, otherwise it cannot substantiate the claim. *See, e.g., Dole v. Milonas*, 889 F.2d 885, 888 n.3 (9th Cir. 1989). An undefined “list serv” is not a “sender” nor “recipient.” Plaintiffs offered to accept written lists of employees in lieu of documents. RFP 126 seeks documents related to Meta’s SRT platform, an internal database allegedly used to obtain legal advice. But SRT appears to be a tool for communicating with counsel for any purpose, including business purposes. Plaintiffs need sufficient information to assess Meta’s anticipated privilege claims over the SRT platform.

## II. META'S STATEMENT

### A. Meta Appropriately Narrowed Plaintiffs' Vague and Overbroad Defined Terms

**“Llama Models:** Meta only recently released the latest versions of Llama 3, and Llama 4 remains under development. Llama 5's development is even more nascent than Llama 4's. Nevertheless, Meta has represented that it is not withholding responsive materials on the ground that they concern Llama 5. Moreover, and as Plaintiffs know, the search terms used by Meta to identify responsive documents are not limited to specific Llama models. Plaintiffs have also been free to ask deponents about Llama 5.

Just a few weeks ago, Plaintiffs moved for discovery into Llama 4, which Meta also was not withholding and had been producing. Dkt. 279. Yet Plaintiffs' ceaseless demands continue, now for models even more distant into the future. They first argue that Llama 5 is relevant because “the class period” runs through the present—but there is no class period because class discovery has not yet begun. January 12, 2024 CMC Hearing Tr. 11:9-13, 12:10-11. Second, they contend that documents concerning Llama 5 may reflect “ongoing” infringement by Meta. But whether Meta eventually trains future models on datasets that allegedly contain Plaintiffs' books (1) will not alter the analysis of whether Meta's use is fair, and (2) can be readily addressed by a limited supplementation sufficient to show that fact if it occurs, without requiring Meta to revisit Plaintiffs' broad discovery in toto.<sup>3</sup> This is all that Rule 26(e)(1)(A) requires. *See Valeo Schalter und Sensoren GmbH v. NVIDIA Corp.*, 2024 WL 3916101, at \*2 (N.D. Cal. Aug. 23, 2024) (duty to supplement is “triggered only if it learns that its prior production is materially incomplete or incorrect” and limited to “‘material’ changes”).

**“Shadow Datasets”:** This term is defined by Plaintiffs as “the type of databases described in paragraph 37 of the Complaint,” which, in turn, refers vaguely to “notorious ‘shadow library’ websites” that “have long been of interest to the AI-training community.” Meta cannot determine which datasets this refers to other than the examples identified by Plaintiffs (which Meta has produced discovery on). Meta has even offered to address other datasets if Plaintiffs identify them, but Plaintiffs have declined to do so.

Plaintiffs insinuate that Meta is using a narrow construction of “Shadow Datasets” to withhold information about Meta's training data, but this is false. Meta has disclosed all datasets that were used for Llama 1, 2, and 3, and has produced numerous documents sufficient to show various other datasets that have been or are being considered for use with future models.

**“Agreements”:** From the outset of discovery, Plaintiffs have defined “Agreements” broadly to encompass, among other things, “any oral... contract, arrangement or understanding, whether formal or informal.” And since the beginning of discovery, Meta has objected to this language. Ex. 1. Plaintiffs do not explain how there will be documents about “oral arrangements” or “understandings,” let alone “informal” ones. Plaintiffs also cite no evidence of such agreements existing (e.g., deposition testimony), nor explain how Meta would produce documents of such agreements.

---

<sup>3</sup> Plaintiffs' burdensome demand for source code for a future, undeveloped model, is especially misplaced, as this case is not about source code and source code is not necessary to (or even the easiest way to) identify training data or other relevant facts. Plaintiffs did not even ask for source code until the final days to serve written discovery. ECF 288 at 4.

Plaintiffs’ purported concern that Meta is using a narrower construction of “Agreements” to avoid producing documents is misplaced. Meta expressly construed “Agreements” to include “drafts” and “versions”; Meta has agreed to produce documents and communications concerning its efforts to license text data, if any; and, as Plaintiffs are well aware, Meta has produced numerous documents and communications in response to recent discovery concerning negotiations that never materialized into agreements. Meta also is not withholding documents concerning oral agreements, if any.

**B. RFP Nos. 114, 115, and 117 Concern Irrelevant Subject Matter**

**Right of Publicity Licenses:** RFP 114 seeks “all” documents and communications concerning agreements between Meta and certain celebrities to use their likenesses, i.e., their publicity rights, for AI chatbots. This subject matter is neither relevant nor proportional to Plaintiffs’ needs. Publicity rights are distinct from copyright, and Meta’s use of celebrity voices to create talking chatbots that recognizably sound like those celebrities has no relevance to the issues in dispute. *See In re NCAA Student-Athlete Name & Likeness Licensing Litig.*, 724 F.3d 1268, 1274 (9th Cir. 2013) (discussing First Amendment limitations on publicity rights and distinguishing from copyright fair use).

**EU Release:** Plaintiffs assert, without explanation or substantiation, that Meta’s decision to withhold release of Llama models from the EU “bears on Meta’s willful infringement and knowledge of the risks of training Llama on pirated works.” As this Court already explained in sustaining Meta’s objections to Topic 5 of Plaintiffs’ Amended 30(b)(6) notice, this subject is “a huge detour into a large subject that has minimal relevance to this case.” ECF 252 at 2. The same result should follow here.

**Public Disclosure of Training Data:** Meta’s decision to refrain from disclosing all training data for its models—a practice that is common among major generative AI developers—is not relevant. Plaintiffs suggest Meta chose not to do so because “the training databases contain pirated versions of copyrighted works,” but this rank, conclusory speculation does not support the relevance of such documents, let alone why Plaintiffs are entitled to “all” documents and communications concerning the same.

**C. Meta’s “Sufficient to Show” Limitations Are Appropriate.**

**Licensing Agreements “for Llama Models”:** RFP 91 concerns “licensing agreements for Llama Models,” i.e., the licenses under which the models were distributed. This subject matter is of ancillary relevance at best and, on that basis, Meta agreed to search for documents sufficient to show such licenses. Plaintiffs attempt to rewrite this Request to concern licensing agreements “for training data for the Llama Models.” But that is plainly not what was requested and, if it were, RFP 91 would be wholly duplicative of RFP 130. Furthermore, as to RFP 130, Meta has conducted a reasonable search for documents and communications concerning its efforts to license text data for the Llama models.

**Financial Data:** RFPs 94 and 100-101 broadly seek “all documents and communications” concerning various financials associated with the Llama models. Insofar as these Requests seek “all” such documents and communications, they are overbroad and disproportionate to Plaintiffs’ needs. Nevertheless, Meta has produced financial documents reflecting costs, revenues, and financial projections associated with the Llama Models to the extent that they exist and could be located following a reasonable search. Meta has also conducted reasonable searches for

communications concerning broad categories of financial information sought by Plaintiffs and is not withholding responsive documents located by its searches. Finally, the Court should reject Plaintiffs’ demand to extend the document requests beyond their written scope to cover documents that are about Meta’s broader Generative AI work unrelated to Llama.

**Training Guidelines:** Meta has produced voluminous documents and communications responsive to RFP 106 and logged countless others that are subject to privilege, in whole or in part. Plaintiffs have also questioned Meta witnesses about the same subject matter, including a 30(b)(6) deponent. Yet Plaintiffs do not identify any deficiency in what was provided, they just want “more” regardless of how duplicative or burdensome it may be. Plaintiffs’ demands should be rejected.

**D. Meta Agreed to Produce Commits and Pull Requests as Text Files, Mooting This Issue**

**E. Plaintiffs’ Discovery on Discovery is Improper**

Plaintiffs concede that the only purported relevance of **RFP 125** is to evaluate certain of Meta’s privilege claims. Because this Court has twice ruled that Plaintiffs’ challenge to Meta’s privilege log was untimely, Plaintiffs’ demand to compel responsive documents should be denied. *See* ECF Nos. 279, 285. It should be denied for the additional reason that RFP 125 improperly constitutes discovery on discovery without a demonstrated need, as the Court has previously denied. ECF 252 at 2.

**RFP 126** is overbroad, unduly burdensome, and untethered to any relevant issue in that it seeks “all documents and communications” related to Meta’s decision to use its “SRT” tool to communicate legal advice. This request should be denied for the same reason as RFP 125 and for the additional reasons that (1) it rests on baseless conjecture that the SRT is used for *any* purpose; and (2) Plaintiffs’ purported need to probe the purpose of SRT is adequately addressed by RFP No. 124, which seeks “Documents or Communications sufficient to show the functionalities and purpose of the SRT” and is not in dispute.

**III. PLAINTIFFS’ REPLY**

In limited page space, Plaintiffs raise these significant points without conceding the other issues.

**Llama 5:** This is not a new ask. *See* Dkt. 267 at 10 n.9. Meta is also wrong to state that this is not a source code case. As Meta knows, expert discovery will involve analysis of source code to determine the extent to which, and how, Llama and other products use copyrighted works.

**Agreements:** The fact that Meta has recurrently raised a meritless definitional objection bears no weight into whether the objection is valid for *these* RFPs. Further, to the extent Meta is not withholding documents reflecting oral agreements (e.g., any emails stating Meta reached agreements with third parties on licensing), it should have so stated instead of contesting the issue.

**Public Disclosure of Training Data:** This RFP is relevant because Meta is unique among LLM developers in that it *did* publicly disclose training data for Llama 1, but then changed course and *did not* disclose that data for all subsequent Llama models. That change of course, on which Meta continues to withhold all discovery, is a core business decision relevant to copyright infringement.

**Sufficient to Show Limitations & RFPs 126-127:** Meta fails to support any assertion of burden (or even identify the burden) as to *any* RFP identified by Plaintiffs. Meta also doesn’t attempt to explain what is supposedly duplicative or beyond the scope as to any financial or training RFP. And Meta ignores entirely that it continues to withhold and redact documents in response to Plaintiffs’ additional discovery requests and has not yet served updated privilege logs



By: /s/ Bobby Ghajar

Bobby A. Ghajar  
Colette A. Ghazarian  
**COOLEY LLP**  
1333 2<sup>nd</sup> Street, Suite 400  
Santa Monica, CA 90401  
Telephone: (310) 883-6400  
Facsimile: (310) 883-6500  
Email: bghajar@cooley.com  
cghazarian@cooley.com

Mark R. Weinstein  
Elizabeth L. Stameshkin  
**COOLEY LLP**  
3175 Hanover Street  
Palo Alto, CA 94304  
Telephone: (650) 843-5000  
Facsimile: (650) 849-7400  
Email: mweinstein@cooley.com  
lstameshkin@cooley.com

Kathleen R. Hartnett  
Judd D. Lauter  
**COOLEY LLP**  
3 Embarcadero Center, 20<sup>th</sup> Floor  
San Francisco, CA 94111  
Telephone: (415) 693-2071  
Facsimile: (415) 693-2222  
Email: khartnett@cooley.com  
jlauter@cooley.com

Phillip Morton  
**COOLEY LLP**  
1299 Pennsylvania Avenue, NW, Suite 700  
Washington, DC 20004  
Telephone: (202) 842-7800  
Facsimile: (202) 842-7899  
Email: pmorton@cooley.com

Angela L. Dunning  
**CLEARY GOTTlieb STEEN &  
HAMILTON LLP**  
1841 Page Mill Road, Suite 250  
Palo Alto, CA 94304  
Telephone: (650) 815-4121

By: /s/ Maxwell V. Pritt

**BOIES SCHILLER FLEXNER LLP**  
David Boies (*pro hac vice*)  
333 Main Street  
Armonk, NY 10504  
(914) 749-8200  
dboies@bsfllp.com

Maxwell V. Pritt (SBN 253155)  
Joshua M. Stein (SBN 298856)  
44 Montgomery Street, 41st Floor  
San Francisco, CA 94104  
(415) 293-6800  
mpritt@bsfllp.com  
jischiller@bsfllp.com  
jstein@bsfllp.com

Jesse Panuccio (*pro hac vice*)  
1401 New York Ave, NW  
Washington, DC 20005  
(202) 237-2727  
jpanuccio@bsfllp.com

Joshua I. Schiller (SBN 330653)  
David L. Simons (*pro hac vice*)  
55 Hudson Yards, 20th Floor  
New York, NY 10001  
(914) 749-8200  
dsimons@bsfllp.com

*Interim Lead Counsel for Plaintiffs*

Facsimile: (650) 849-7400  
Email: [adunning@cgsh.com](mailto:adunning@cgsh.com)

*Attorneys for Defendant Meta Platforms, Inc.*



**ATTESTATION PURSUANT TO CIVIL LOCAL RULE 5-1(h)**

I hereby attest that I obtained concurrence in the filing of this document from each of the other signatories. I declare under penalty of perjury that the foregoing is true and correct.

Dated: November 26, 2024

BOIES SCHILLER FLEXNER LLP

/s/ Maxwell V. Pritt

Maxwell V. Pritt

Reed Forbush

Jay Schuffenhauer

Attorneys for Plaintiffs